ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

(Laboratorio di)
Amministrazione di sistemi

Sysadm nell'era del Cloud computing

Marco Prandini

Dipartimento di Informatica – Scienza e Ingegneria

L’evoluzione cloud computing

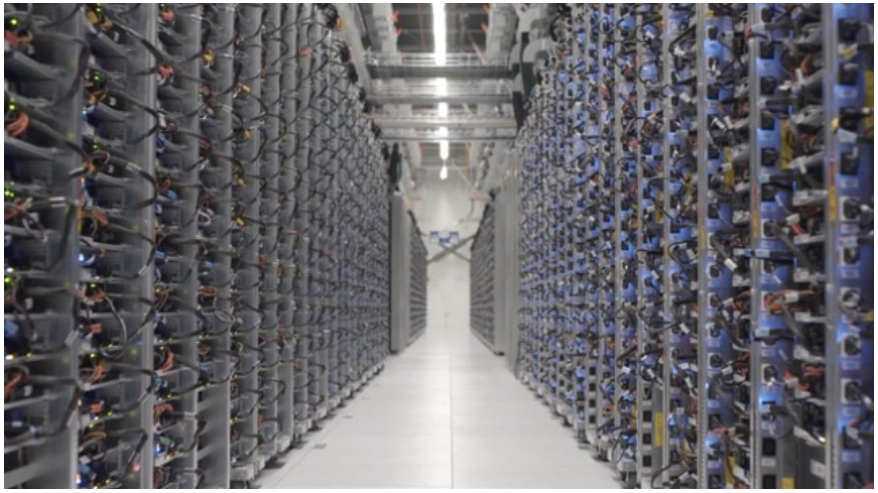
- Le attività “tradizionali” vengono svolte *on premises* – l’infrastruttura cioè è di proprietà e direttamente accessibile → totalità del controllo, ma anche dei problemi connessi a tutti gli strati che non sono “core”
- Tra i problemi più gravosi, e più raramente di interesse diretto delle aziende che hanno necessità di sistemi di elaborazione, ci sono
  - Alta disponibilità
  - Sicurezza
  - Aggiornamento hardware

Ambiente

- Ovvietà: per erogare con continuità un servizio, il sistema deve essere acceso e connesso!
  - La collocazione realmente *on premises* è sempre più rara
  - Riservata a casi in cui
    - la complessità gestionale di esternalizzare non è giustificata
    - non ci si può permettere di dipendere da fattori esterni per raggiungere i servizi erogati
    - non si può delegare il controllo fisico dei dati o rischiare di essere dipendenti dalle scelte architetturali di altri

Ambiente

- Esternalizzare significa affidare a professionisti la gestione di un ambiente affidabile
  - Le strutture per realizzare tale ambiente sono complesse e molto costose → indispensabile condividerle



Ambiente

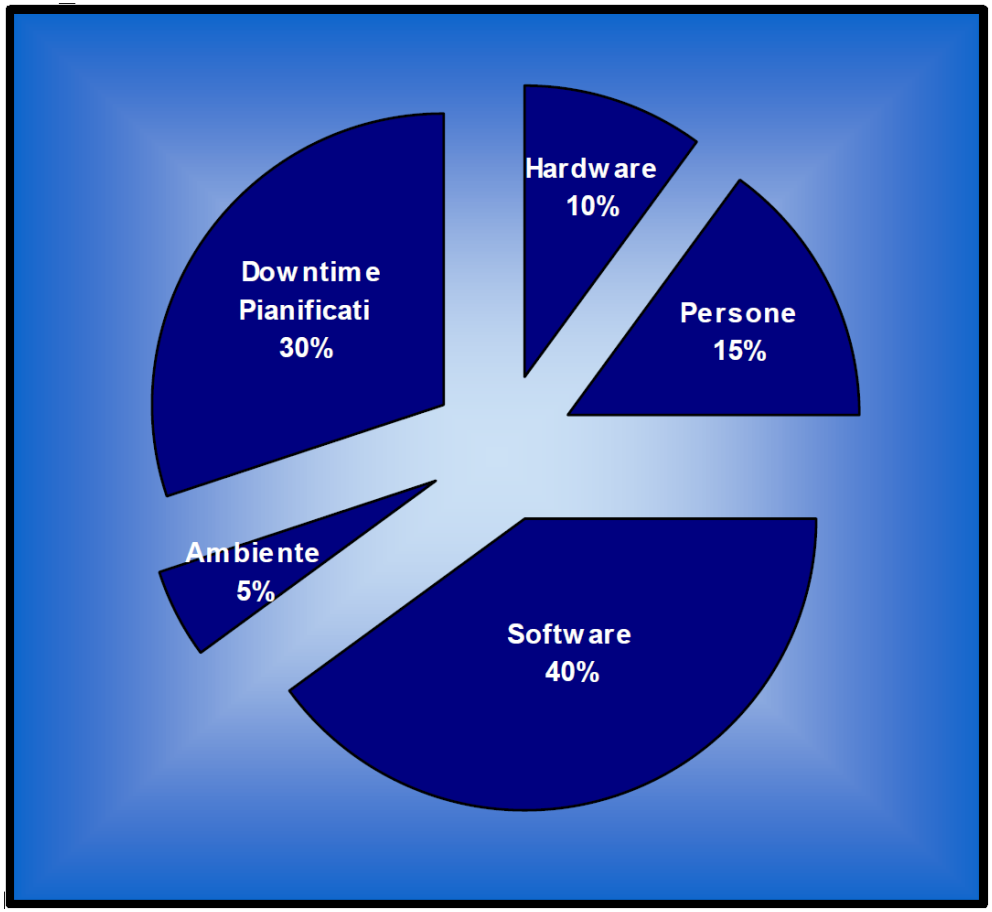
https://datacenter.com/news_and_insight/data-center-redundancy-2plus1-2n-distributed-redundancy/

- *Data center* o *server farm* sono i luoghi in cui vengono ospitati in grande quantità i sistemi di calcolo
  - Housing o Co-location: fornitura di spazio e connettività per sistemi acquistati e gestiti dal cliente
  - Managed housing: fornitura dei sistemi in housing (su hardware comunque dedicato al cliente) e loro gestione sistemistica
  - Hosting: fornitura di uno o più servizi specifici (storage, web, posta, …) su hardware condiviso tra più clienti
  - **Il modello cloud è un caso speciale dell'hosting tradizionale**

Disponibilità / Downtime

- Esternalizzando in tutto o in parte la collocazione, servono garanzie
- La *disponibilità* o *availability* di un sistema è l'indicatore più semplice: il rapporto tra il tempo per cui eroga correttamente i servizi (***u**ptime*) rispetto al tempo per cui ci si attende che lo faccia (tempo di **o**sservazione)

  A=U/O
- Il tempo durante il quale il servizio non è erogabile viene chiamato *downtime*



Cause di downtime per incidenza sul totale

Livelli di disponibilità

- Comunemente la disponibilità viene indicata in modo sintetico col "numero di 9" nella percentuale di uptime
  - aggiungere un 9 significa dividere per 10 il downtime

disponibilità %	downtime per anno	downtime per mese	downtime per settimana
98%	7,3 giorni	14,4 ore	3,36 ore
99%	3,65 giorni	7,20 ore	1,68 ore
99,5%	1,83 giorni	3,60 ore	50,4 minuti
99,9%	8,76 ore	43,2 minuti	10,1 minuti
99,99%	52,6 minuti	4,32 minuti	1,01 minuti
99,999%	**5,26 minuti**	**25,9 secondi**	**6,05 secondi**
99,9999%	31,5 secondi	2,59 secondi	0,605 secondi

Service Level Agreement

- L'uptime è solo uno degli aspetti
- Un contratto sul livello di servizio (*Service Level Agreement*) può prevedere vincoli più stringenti
  - Sulla distribuzione del downtime (es.: "four nines" nell'arco dell'anno ma in frazioni non superiori a *n* minuti consecutivi)
  - Su parametri di servizio non collegati al downtime
    - Prestazioni dei servizi
    - Tipologie di assistenza previste e loro caratteristiche
    - ...
- Oggetto del contratto tipicamente è anche il contratto stesso
  - Modalità di aggiornamento
  - Reportistica periodica sulle variabili monitorate

HA fisica

- Resistenza della struttura
  - cause naturali: terremoti, inondazioni, ...
  - cause artificiali: incidenti aerei e ferroviari, terrorismo, ...

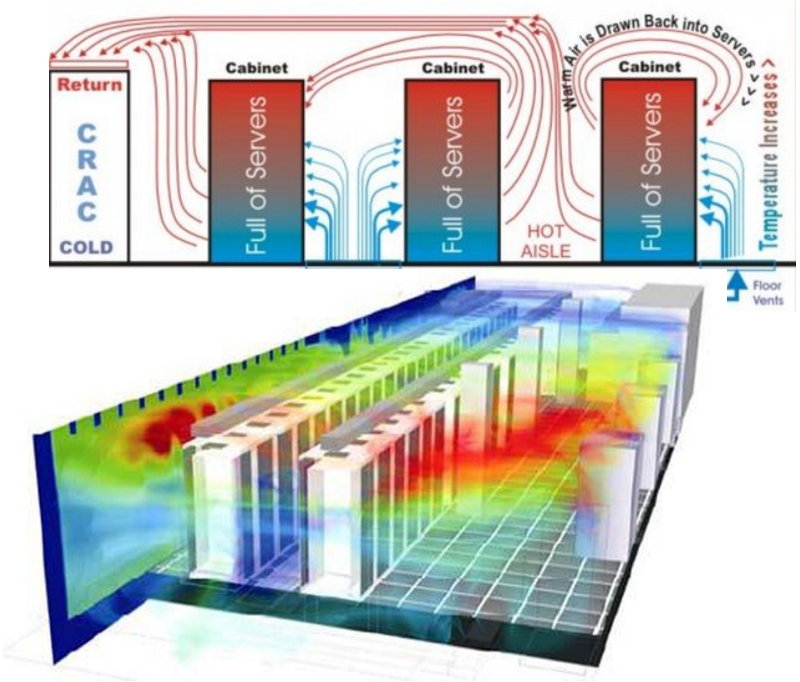
    https://goo.gl/maps/zUwqeZJrJrQU7eG97
  - cause interne: incendi, da controllare con sistemi che consentano l'intervento anche quando l'incendio stesso li danneggia parzialmente
    - ogni sistema complesso, anche se introdotto per limitare danni, può causarne altri di imprevisti

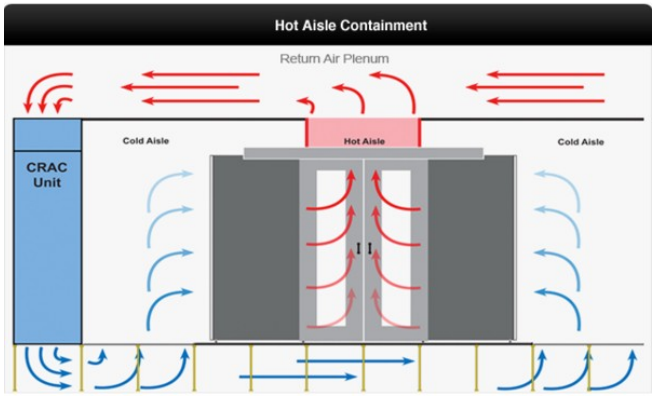
      https://journal.uptimeinstitute.com/fire-suppression-systems-bring-risk/
- Sicurezza e controllo degli accessi
  - perimetro esterno blindato e sorvegliato da staff (armato) 24/7
  - segmentazione settori con liste di controllo accessi separate e concordate in anticipo sull'ingresso
  - apertura varchi a più fattori
  - videosorveglianza con registrazione off-site

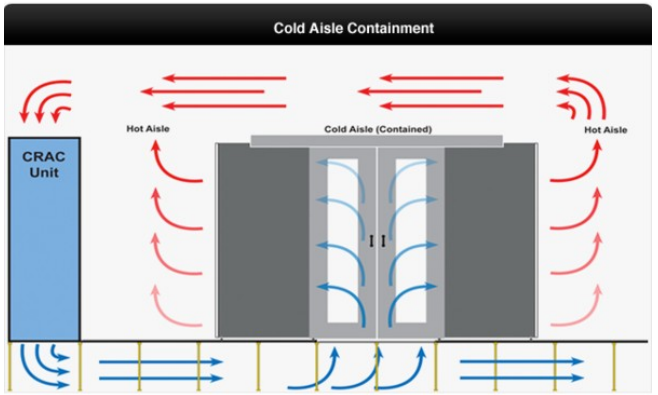
HA ambientale

- Condizionamento dell'aria
  - gestione di temperatura e umidità con sistemi tolleranti ai guasti e alle interruzioni di erogazione dell'energia elettrica

https://www.colocationamerica.com/blog/cooling-innovations-for-data-centers



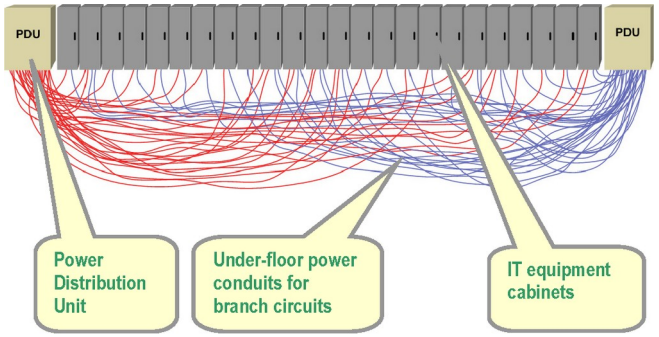




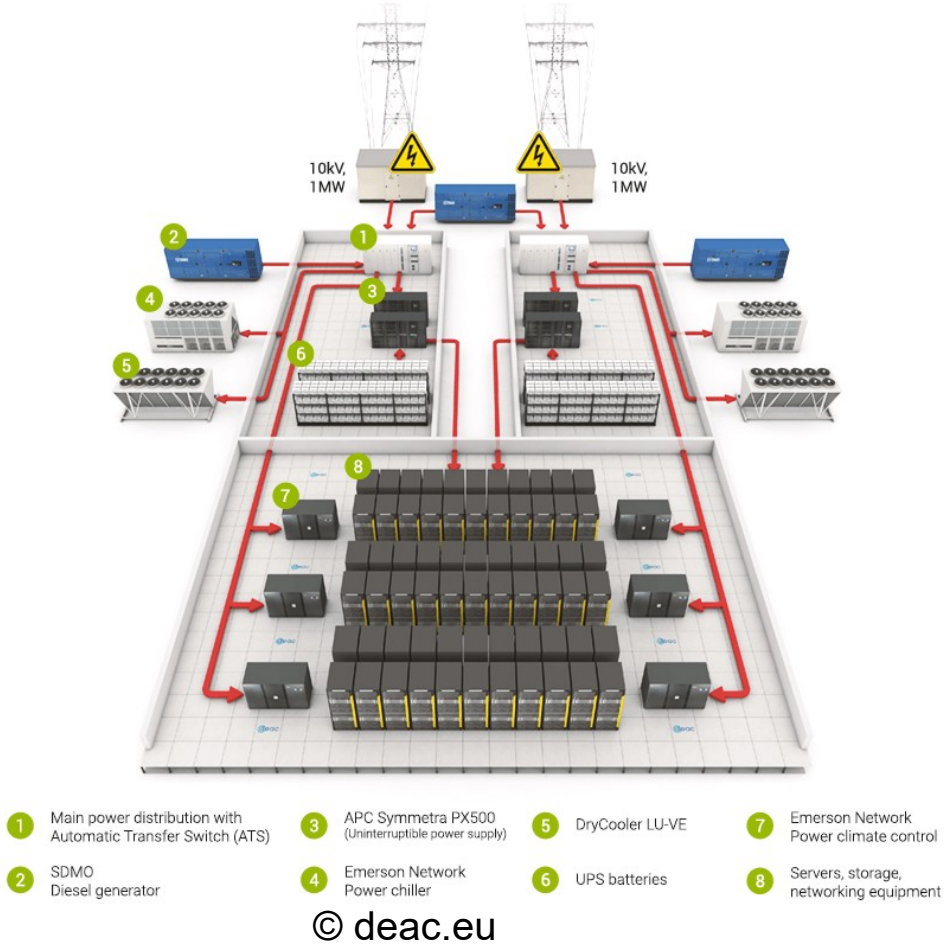
https://datacenterresources.com/articles/identifying-data-center-cooling-issues/

HA infrastrutturale

- Condizionamento dell'alimentazione elettrica
  - erogazione su almeno due linee indipendenti per ogni apparato



  - pulizia della sinusoide per allungare la vita degli apparati
  - sistemi di continuità ad intervento istantaneo e di durata prolungata
    - motogeneratori → lunga durata, avviamento lento
    - batterie → intervento istantaneo, bassa capacità



- Gestione dei consumi

https://www.thegreengrid.org/en/resources/library-and-tools/20-PUE:-A-Comprehensive-Examination-of-the-Metric

HA della connettività

- Connettività di rete
  - connessione tramite provider indipendenti
    - è comune per i datacenter principali avvalersi di oltre 10 carrier
    - spesso fungono da internet exchange

https://www.datacentermap.com/singapore/singapore/equinix-singapore_connectivity.html

https://www.datacentermap.com/usa/california/los-angeles/one-wilshire_connectivity.html

https://www.datacentermap.com/united-kingdom/london/telehouse-london-north_connectivity.html

  - collocazione fisica dei cavi su percorsi indipendenti

https://www.datacenterdynamics.com/en/news/google-cloud-us-east1-data-centers-disrupted-due-physical-damage-multiple-fiber-bundles/

Business continuity

- Per garantire l'accessibilità senza interruzioni di dati e servizi sono state sviluppate alcune tecniche di base che partono dagli elementi costruttivi e salgono alle architetture
  - robustezza dei supporti di storage → RAID
  - flessibilità di allocazione dello storage → LVM
  - robustezza nell'accesso allo storage → multipath
  - ridondanza dei nodi di elaborazione → clustering
- Queste tecniche non difendono da eventi
  - improbabili ma catastrofici (cataclismi, attentati, …)
  - limitati ma frequenti (errori degli operatori, attacchi)

  che minano l'accessibilità a lungo termine dei dati
  - → backup, piani di disaster recovery

Basta portare le macchine in datacenter?

- Una semplice soluzione di housing risolve molti problemi di HA ma lascia comunque nelle mani del cliente aspetti sistemistici che spesso non sono di suo interesse diretto
- Esempio:
  - Se il mio business è sviluppare e testare un applicativo desktop, l'hardware non mi interessa, mi basta poter configurare liberamente i diversi sistemi operativi su cui distribuirlo
  - Se è gestire una piattaforma di servizi via web, mi basta poter caricare applicazioni su di un server che le esegua
  - Se è gestire una rete commerciale, mi basta poter caricare e interrogare i database, le agende, i listini, ecc.

L’evoluzione cloud computing

- Oggigiorno, il *cloud computing* permette di affrontare con maggior efficienza molte tipologie di progetti
- Dal punto di vista economico, permette investimenti più flessibili
  - per progetti piccoli o con fattori di utilizzo previsto lontani dal 100%
    - intendendo l'utilizzo medio rispetto alla capacità di picco sulla quale verrebbe dimensionato l'acquisto
    - caso tipico: workload fortemente stagionali o concentrati in ore del giorno
  - per progetti medi e grandi al punto da rendere difficoltoso il forte investimento in conto capitale
  - per progetti con aspettative di forte crescita, ma senza certezze dei tempi in cui si concretizzerà
- Dal punto di vista gestionale, esternalizza le attività non-core

Cloud computing: concetti base

- Un cloud provider si fa carico della realizzazione di un (gruppo di) data center allo stato dell'arte
  - realizza gli edifici
  - predispone gli impianti
  - acquista ingenti quantità di apparati di calcolo e networking di diverse fasce
- Il pool di risorse complessive viene utilizzato per far funzionare sistemi virtualizzati
  - molti clienti condividono le risorse fisiche (multi-tenancy) spalmando i costi fissi e delegando completamente la loro amministrazione
  - la configurazione è tramite interfacce che nascondono completamente la struttura fisica
  - provisioning dinamico: l'avvio e arresto delle risorse è *on demand*
    - il pagamento è solo per il periodo di effettivo utilizzo
  - scalabilità: la dimensione del provider tipicamente dà l'illusione al cliente di poter allocare illimitatamente nuove risorse al bisogno

→ RISORSE "As A Service"

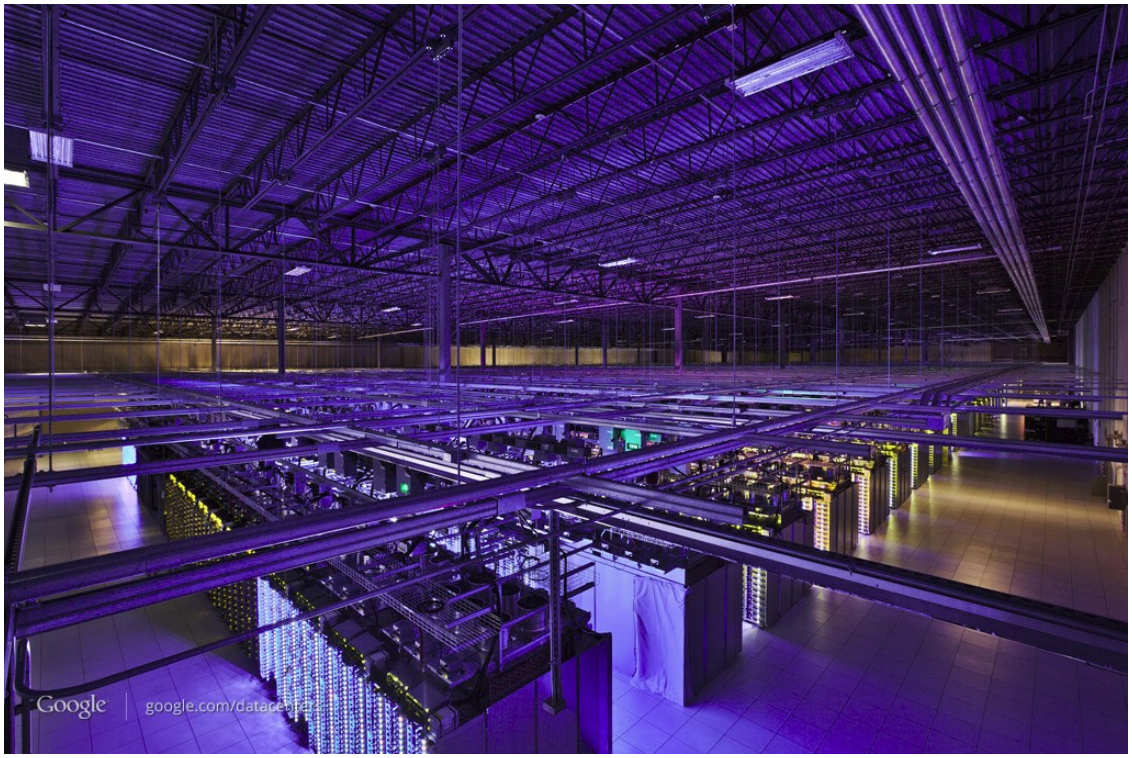
Livelli e attori

tratto da Principles, Applications and Models for Distributed Systems M – Antonio Corradi & Luca Foschini

- *aaS = Everything as a Service
- SaaS – Software as a Service
  - Le risorse sono applicazioni rese disponibili via web agli utenti
    - Gmail, Dropbox, SalesForce, Teams, …
- PaaS – Platform as a Service
  - Le risorse sono intere piattaforme disponibili per l'esecuzione remota di codice caricato dall'utente
    - web hosting con vari linguaggi server side, cms estendibili, …
- IaaS – Infrastructure as a Service
  - Le risorse sono componenti architetturali virtualizzate
    - hardware per calcolo
    - sistemi operativi
    - dispositivi di networking

Livelli e attori

tratto da Principles, Applications and Models for Distributed Systems M – Antonio Corradi & Luca Foschini

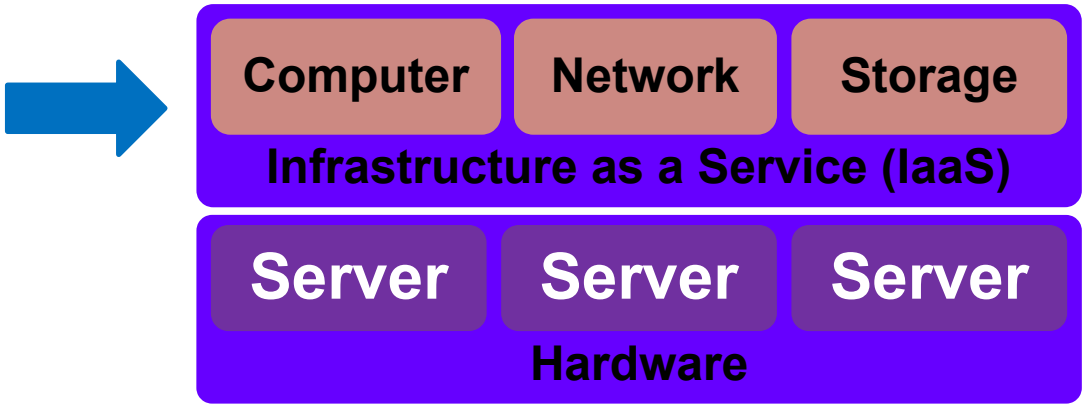


- alla base di tutto, l'architettura reale

Livelli e attori

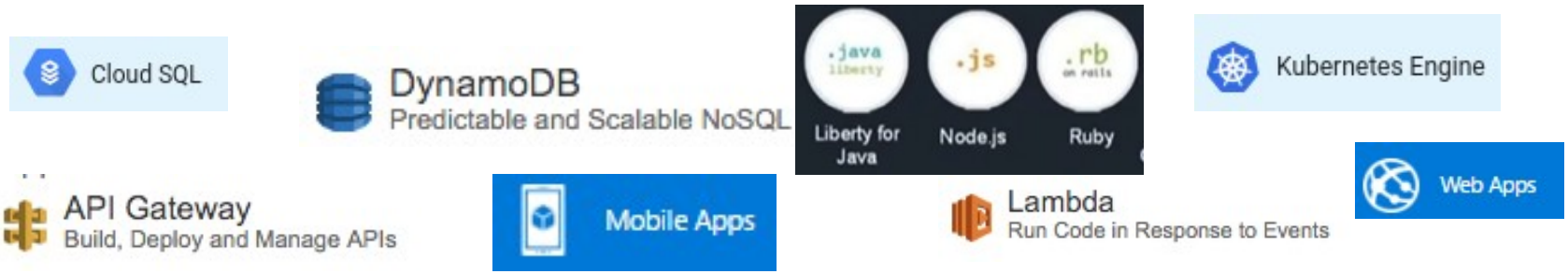
tratto da Principles, Applications and Models for Distributed Systems M – Antonio Corradi & Luca Foschini

- Lo strato infrastrutturale abilita la realizzazione dei servizi cloud, per mezzo della gestione della virtualizzazione
- Si possono ottenere on demand capacità di calcolo, memoria e comunicazione, che poi vanno gestite come se fossero di proprietà

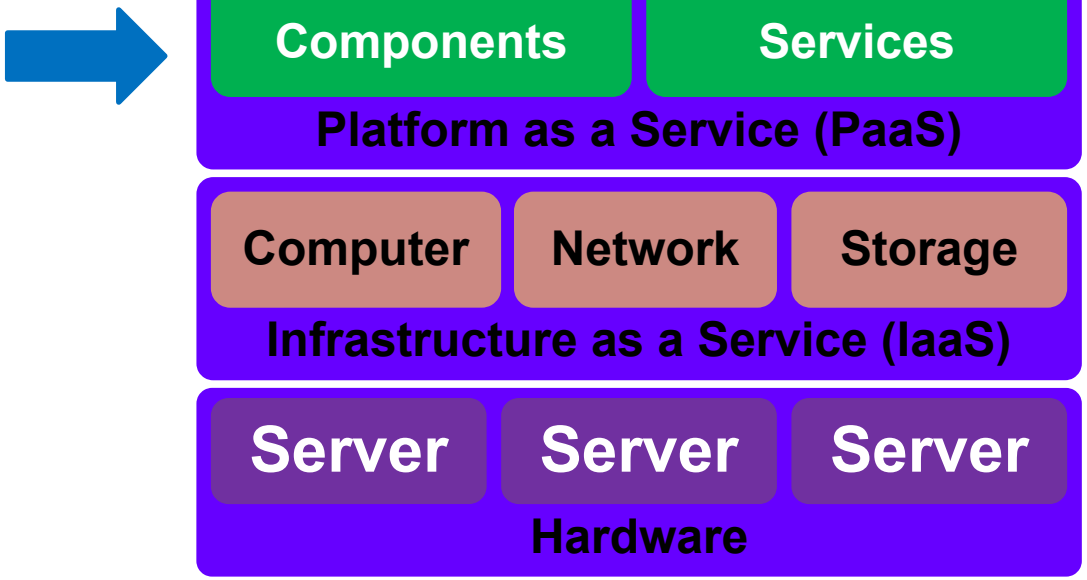


Livelli e attori

tratto da Principles, Applications and Models for Distributed Systems M – Antonio Corradi & Luca Foschini



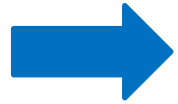
- Lo strato di piattaforma fornisce servizi standard e componenti modulari fruibili da remoto agli strati superiori
- Si evita di gestire l'intero stack sistemistico, e si scrive la logica delle applicazioni

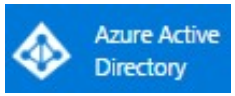


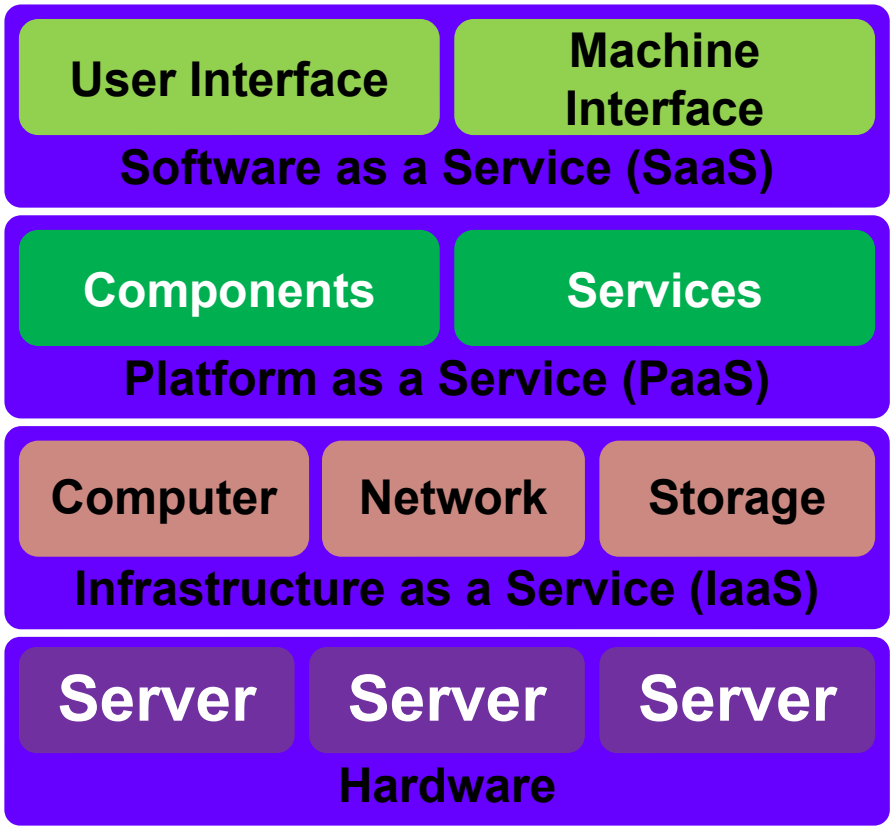
Livelli e attori

tratto da Principles, Applications and Models for Distributed Systems M – Antonio Corradi & Luca Foschini

- Lo strato software mette a disposizione applicazioni preinstallate a cui fornire solamente configurazione e dati





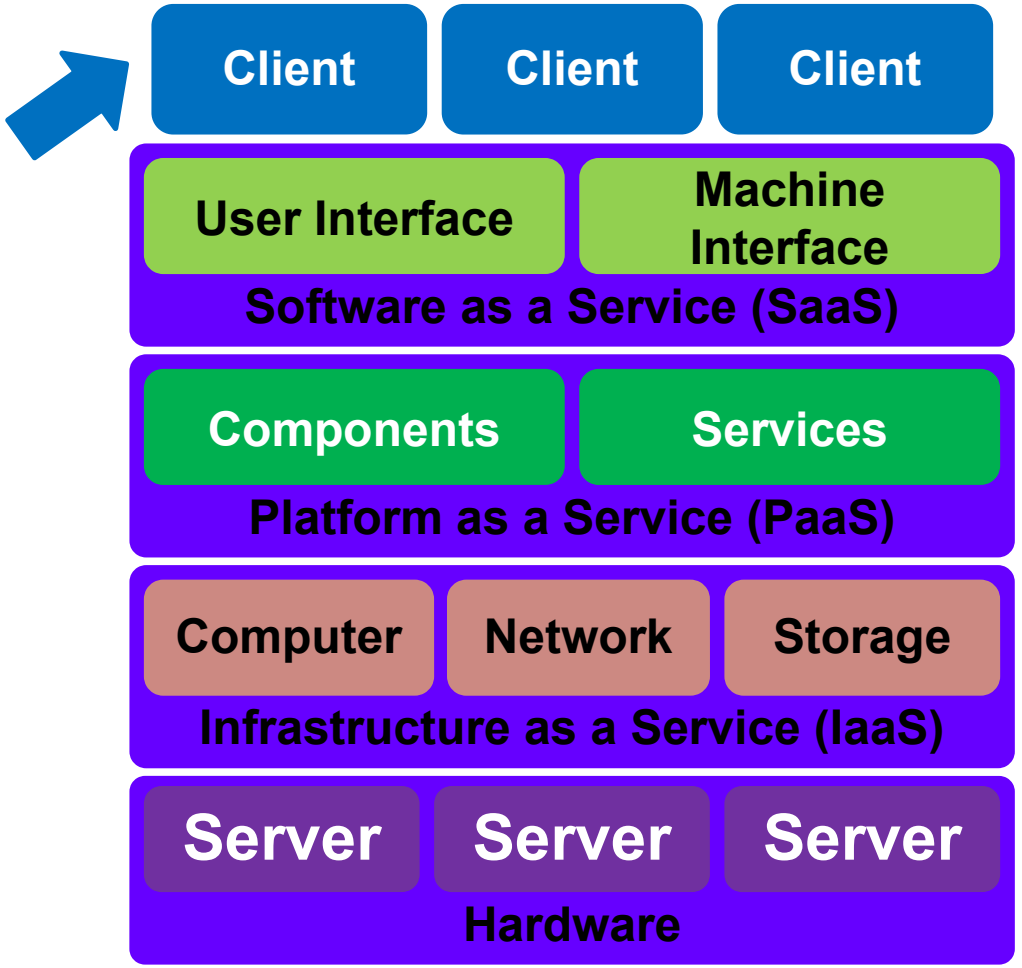


Livelli e attori

tratto da Principles, Applications and Models for Distributed Systems M – Antonio Corradi & Luca Foschini

- I client permettono di accedere al cloud.
  Restano l'unico componente in esecuzione sulle piattaforme fisicamente in mano all'utente, che attraverso questi può comunicare con
  - applicazioni
  - sistemi di deploy sulle piattaforme
  - sistemi di configurazione e monitoraggio delle infrastrutture

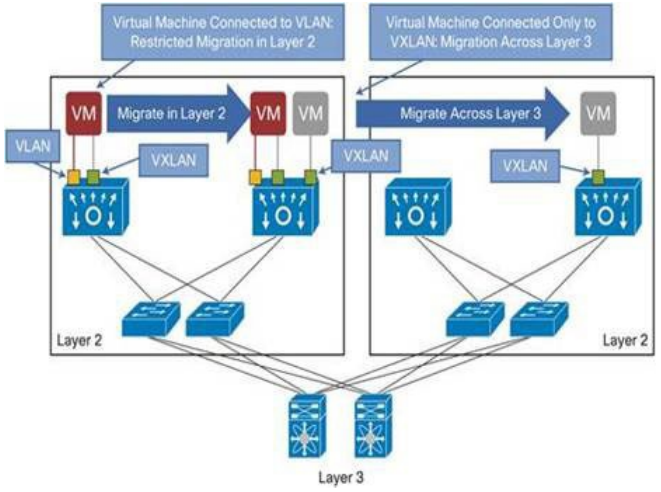
  attraverso i diversi tipi di interfaccia disponibili
  - API
  - Web GUI

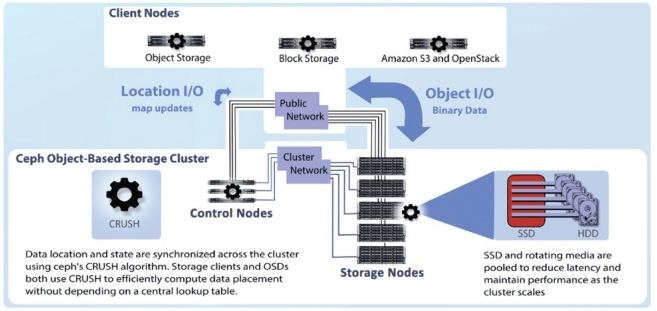


Cloud computing: prerequisiti

Virtualizzazione, virtualizzazione, virtualizzazione

- grandi pool di calcolatori
  - architetturalmente simili
  - intercambiabili
  - su cui gira un hypervisor
- apparati di rete gestibili e riconfigurabili
  - utilizzo massiccio di VLAN per partizionare il traffico tenant
  - vxlan per estendere il layer fisico su scala geografica
  - evoluzione verso Software Defined Networking
- sistemi di storage di rete
  - gerarchici (prestazioni vs costo)
  - ad alta scalabilità





Cloud computing: prerequisiti

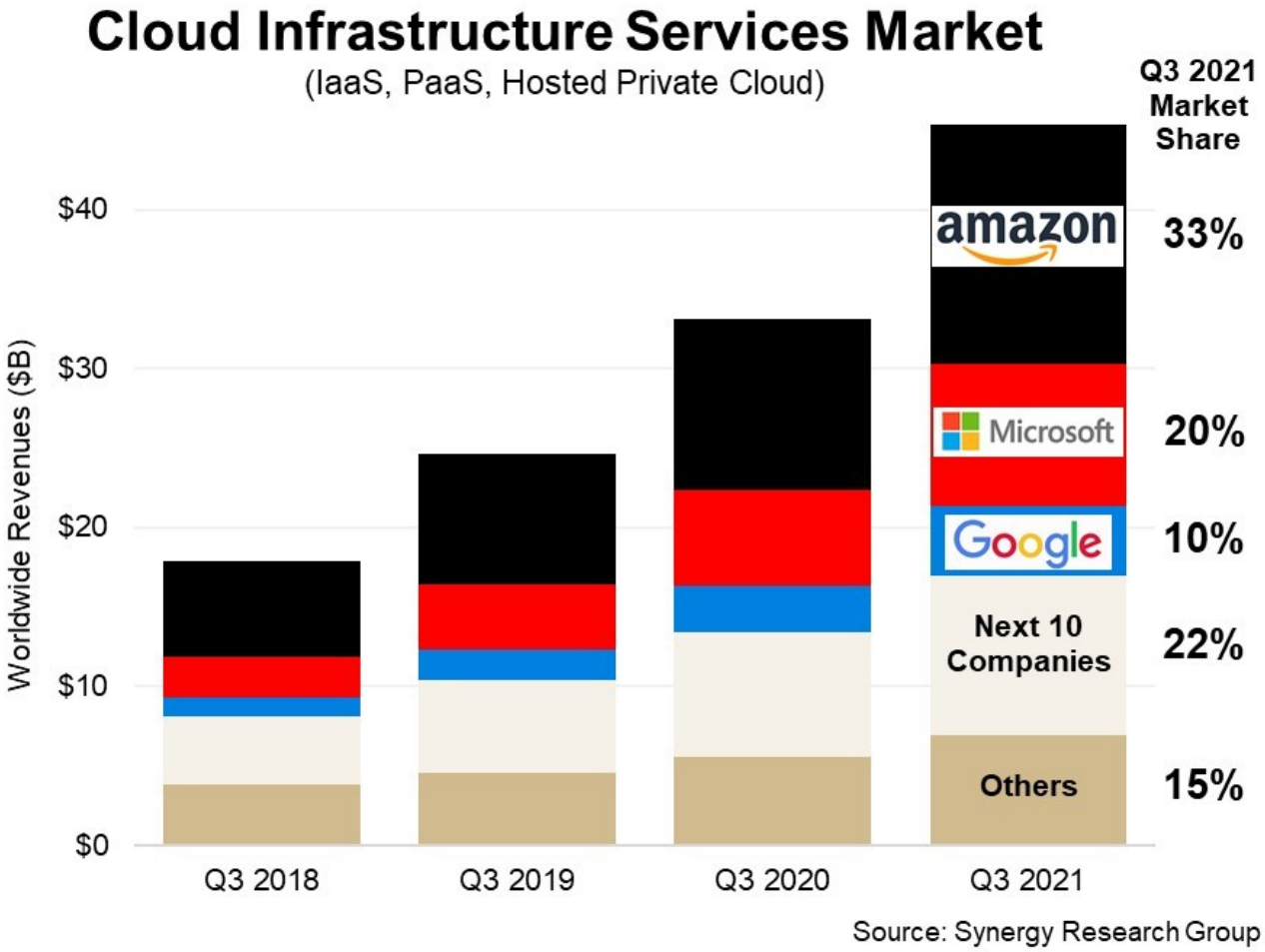
Gestione, gestione, gestione

- interfacce al sistema
  - manuali via web
  - command line
  - integrabili in piattaforme software via API
- sistemi di monitoraggio
  - dettagliati e facilmente accessibili
  - fortemente programmabili per reagire automaticamente a eventi
- modelli di configuration management

    in un ambiente in cui i nodi di erogazione dei servizi formano un pool scalabile, non è più sufficiente saper intervenire sulla configurazione di un servizio, è necessario garantire modifiche coerenti a servizi interdipendenti e propagazione delle modifiche sulle molteplici istanze in esecuzione

  - distribuzione di parametri di configurazione
    - limitato ad aggiornamenti semplici e per i quali è necessario un effetto immediato
  - versioning e templating di file di configurazione
    - **configuration as code**
  - immagini immutabili
    - test → template → sostituzione graduale

Cloud computing: i protagonisti

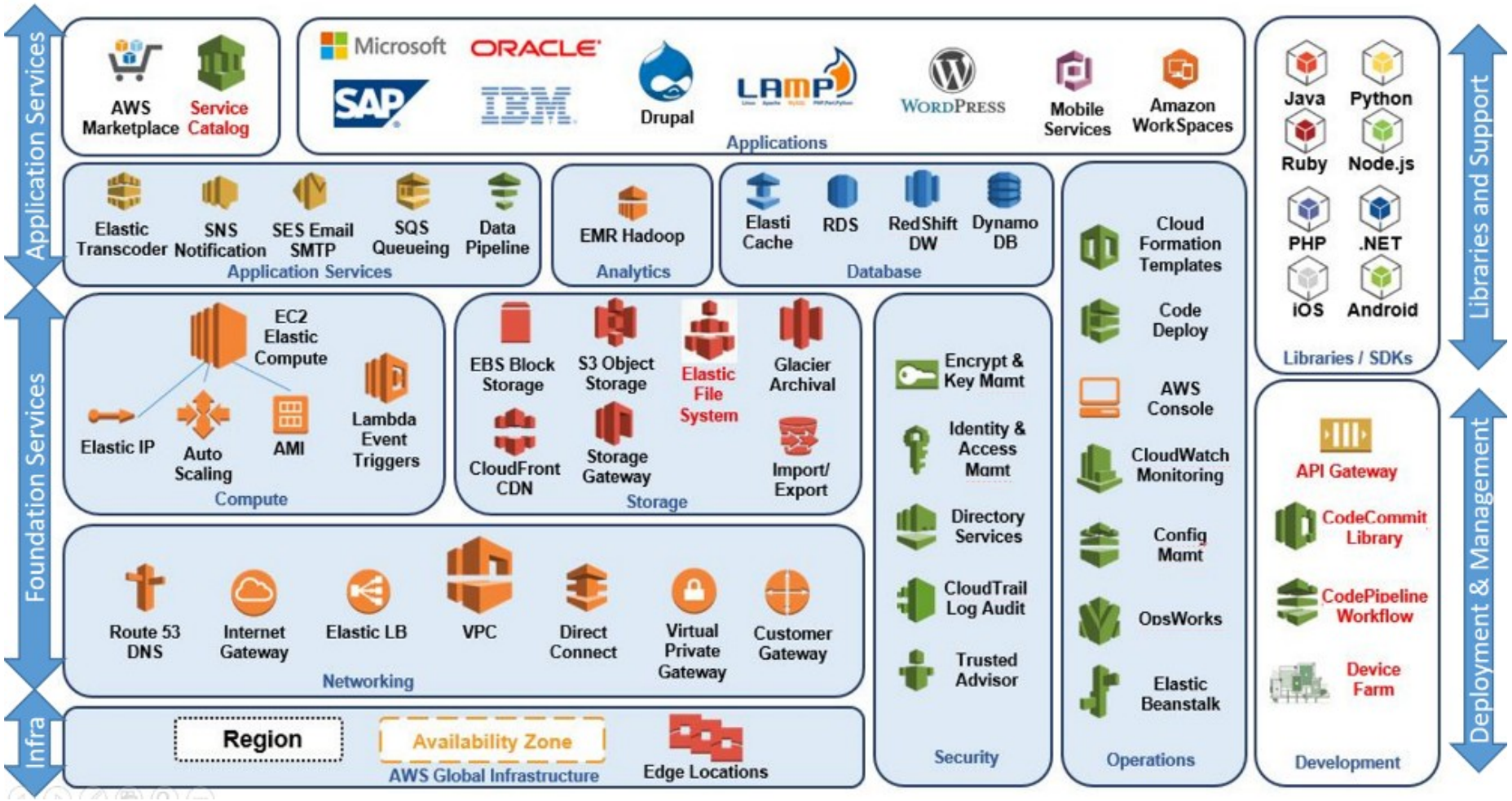


https://www.srgresearch.com/articles/amazon-microsoft-google-grab-the-big-numbers-but-rest-of-cloud-market-still-grows-by-27

Gli strumenti

- Necessità di utilizzare strumenti per gestire le risorse in modo riproducibile, anziché configurarle “artigianalmente”
- DevOps: convergenza tra i team tradizionalmente conflittuali degli sviluppatori e dei sistemisti

Da sysadm ad architetto



http://awsprocert.com/aws-managed-services/